Thailand Statistician
January 2026; 24(1): 12-25
http://statassoc.or.th
Contributed paper

Interval Estimation in Location-Scale Family, Using Information Measures

Saeid Maadani, Gholam Reza Mohtashami Borzadaran[*] and Abdolhamid Rezaei Roknabadi
Department of Statistics, Ferdowsi University of Mashhad, Iran
[*]Corresponding author; e-mail: grmohtashami@um.ac.ir

Received: 10 March 2021
Revised: 15 September 2022
Accepted: 1 October 2022

Abstract

Interval estimation is one of the important concepts in statistics. There are different methods to find a confidence interval for distribution parameters, but finding an optimal confidence interval is always a concern. On the other hand, it is important to find a confidence interval for the Shannon information of the ith order statistic when the distribution parameters are unknown. In the present paper, using information measures, we introduce an asymptotic confidence interval for scale parameter and the Shannon information of the ith order statistic related to the location-scale family. Via simulation study, we study the coverage probability and length of confidence intervals. Finally, some examples for hypothesis testing related to the Shannon information of the ith order statistic and the scale parameter are provided.

Keywords: Asymptotic confidence interval, location and scale family, Shannon entropy, varentropy.

1. Introduction

One of the important information measure is the Shannon (1948) entropy. This measure for discrete random variable X with probability mass function $p(x)$ is defined as

$$H(X) = -\sum_{S} p(x) \log p(x), \tag{1}$$

where S is the support of X. A generalization of (1) for a continuous random variable X with the name of differential entropy is defined as

$$h(X) = -\int_{S} f(x) \log f(x) dx, \tag{2}$$

where $f(x)$ and S are the density function and support of X, respectively. Both (1) and (2) are expectation of the random variables $-\log p(X)$ and $-\log f(X)$, which are called information content. The variance of information content also has interesting properties and applications. In computer sciences, $\mathrm{Var}(-\log p(X))$ is taken into consideration with the name of varentropy. The varentropy is a key parameter for estimating the performance of optimal coding, determining the dispersion of sources and capacity of channels, and so on. To study more, one can refer to references such as Bobkov and Madiman (2011), Kontoyannis and Verdu (2014), and Arikan (2016). There are many studies on the applications of the Shannon entropy in statistical sciences, but there are few studies on the varentropy that can be mentioned as follows:

Song (2001) studied the $\mathrm{Var}(-\log f(X))$ for the continuous random variable X and introduced a new intrinsic measure of distributions shape that can be a good alternative to the kurtosis measure. He showed that the $\mathrm{Var}(-\log f(X))$ gives us information similar to the traditional kurtosis measure $\frac{\mu_4}{\sigma^4}$, where μ_4 and σ^2 are the fourth central moment and variance of X, respectively. Following the work of Song (2001), other studies were also done in this regard, such as Zografos (2008) and Enomoto et al. (2013). Also, more recently, in addition to the above, the application of the varentropy in reliability engineering has been conducted by Di Crescenzo and Paolillo (2020), for residual lifetime random variables. Also, a new generalized varentropy based on Tsallis (1988) entropy was introduced by Maadani et al. (2020).

The study of hypothesis testing and goodness of fit testing using the Shannon information has been considered in recent years. To read more, we can refer to the articles of Menendez (2000), Taufer (2002), Pasha et al. (2004), Jager and Wellner (2007), Zamanzade and Arghami (2012) and Afhami et al. (2015). In this article, we use the entropy and important property of the varentropy for obtaining an asymptotic confidence interval for the scale parameter in location-scale family. Furthermore, we will introduce an asymptotic confidence interval for the measure of uncertainty related to the ith order statistics in this family, which can be useful for system designers. The presented confidence intervals have been studied by a simulation method, and finally, some statistical hypothesis tests have been presented in this regard.

One of the important properties of the varentropy, which is of great interest to us in this article, is that the varentropy is an affine invariant measure.

For a continuous random variable X, the varentropy is

$$VE(X) = \mathrm{Var}(-\log f(X)) = E(\log f(X) - E(\log f(X)))^2, \tag{3}$$

where $VE(X)$ is called the varentropy of X.

Suppose that $Y = g(X)$ is a monotone differentiable function of X. Then it can be shown that

$$f_Y(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}.$$

Therefore,

$$VE(g(X)) = \mathrm{Var}(\log \frac{f(X)}{g'(X)}), \;\; h((g(X)) = -\mathrm{E}(\log \frac{f(X)}{g'(X)}).$$

Hence $VE(g(X)) = VE(X)$, if the $g(X) = aX + b$. Therefore, $VE(aX + b) = VE(X)$ and $h(aX + b) = h(X) + log|a|$. In mathematical statistics, a location-scale family is a family of probability distributions parametrized by a location parameter $\mu \in R$, a nonnegative scale parameter $\sigma > 0$, and the density function as $g(x|\mu, \sigma) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$, where $f(x)$ is any density function of random variable X. Examples of location-scale families are exponential, normal, Cauchy, logistic and double exponential (or Laplace) distributions. Because the varentropy is the location and scale-invariant measure, that is independent of the distribution parameters in this family. However, the Shannon entropy is just independent of the location parameter and dependent on the scale parameter.

Table 1 compares the entropy and varentropy for some distributions in the location-scale family. We can see that in all distributions of this family, the varentropy has a certain value, and when we are unaware of the distribution parameters, the varentropy is known, and we will not need to estimate it. Therefore, the error of estimation of variance in the pivotal quantity is zero, and we get a better asymptotic estimate than when the variance is unknown.

Table 1 Comparison of the entropy and varentropy for some distributions

Distribution	Density	Entropy	Varentropy
Uniform	$f(x)=\dfrac{1}{b-a}\ b>a$	$\log(b-a)$	0
Normal	$f(x)=\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	$\dfrac{1}{2}\log(2\pi e\sigma^2)$	$\dfrac{1}{2}$
Exponential	$f(x)=\theta e^{-\theta x}\ \ x>0,\theta>0$	$1-\log\theta$	1
Cauchy	$f(x)=\dfrac{1}{\pi\sigma[1+(\frac{x-\mu}{\sigma})^2]}$	$\log(4\pi\sigma)$	$\dfrac{\pi^2}{3}$
Logistic	$f(x)=\dfrac{\pi}{\sigma\sqrt{3}}\times\dfrac{e^{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}}}{(1+e^{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}})^2}$	$2-\log\dfrac{\pi}{\sigma\sqrt{3}}$	$4-\dfrac{\pi^2}{3}$
Laplace	$f(x)=\dfrac{1}{2\sigma}e^{-\frac{\vert x-\mu\vert}{\sigma}}$	$1+\log(2\sigma)$	1

In the next section, we introduce an asymptotic confidence interval for entropy and the scale parameter in this family.

2. Asymptotic Confidence Interval

It is well known that the central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution.

Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed observations with density function $f(x,\boldsymbol{\theta})$, that $\overline{X}$ is a mean of the sample, and that μ and σ^2 are the mean and variance of X, respectively. The CLT implies that $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Similar to the said above, we consider random samples:

$$Y_1 = -\log f(X_1|\boldsymbol{\theta}), Y_2 = -\log f(X_2|\boldsymbol{\theta}), \ldots, Y_n = -\log f(X_n|\boldsymbol{\theta}).$$

It is easy to show that

$$E(\overline{Y}) = h(X) \text{ and } \mathrm{Var}(\overline{Y}) = \frac{\mathrm{Var}(-\log f(X|\boldsymbol{\theta}))}{n} = \frac{VE(X)}{n}.$$

Then using the CLT we have

$$\frac{\overline{Y}-h(X)}{\sqrt{\frac{VE(X)}{n}}} \sim N(0,1), \tag{4}$$

where $\overline{Y} = -\frac{1}{n}\sum_{i=1}^{n}\log f(X_i,\boldsymbol{\theta})$ and (4) is an asymptotic pivotal variable for our confidence interval. Suppose that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}=(\mu,\sigma)$. Then

$$lim_{n\to\infty}p_r(|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}|>\varepsilon)=0 \text{ for all } \varepsilon>0.$$

Using Theorem 5-5-4 in Casella and berger (1990), for a continuous density function $f(x,\theta)$ and for all $\varepsilon>0$, we have

$$lim_{n\to\infty}p_r(|-\frac{1}{n}\sum_{i=1}^{n}\log f(X_i,\hat{\boldsymbol{\theta}})+\frac{1}{n}\sum_{i=1}^{n}\log f(X_i,\boldsymbol{\theta})|>\varepsilon)=0.$$

Therefore,

$$Z_{\hat{\boldsymbol{\theta}}} = \frac{-\dfrac{1}{n}\sum_{i=1}^{n}\log f(X_i,\hat{\boldsymbol{\theta}})-h(X)}{\sqrt{\frac{VE(X)}{n}}} \tag{5}$$

is convergence to (4) in probability. In the location and scale family with parameter $\boldsymbol{\theta} = (\mu, \sigma)$, the varentropy is a certain value, and the Shannon entropy depends on the scale parameter σ (See Table 1). Therefore, when the parameters of distribution are unknown, the varentropy is known. If $VE(X) = \nu^2$, using (5), then a simple asymptotic confidence interval of the Shannon entropy is

$$(L_{\hat{\boldsymbol{\theta}}} - Z_{\frac{\alpha}{2}} \frac{\nu}{\sqrt{n}}, L_{\hat{\boldsymbol{\theta}}} + Z_{\frac{\alpha}{2}} \frac{\nu}{\sqrt{n}}), \tag{6}$$

where $L_{\hat{\boldsymbol{\theta}}} = -\frac{1}{n} \sum_{i=1}^{n} \log f(X_i, \hat{\boldsymbol{\theta}})$. It is to mention that the varentropy no need to estimate in this family and the known variance is a privilege for calculating a more accurate confidence interval. Therefore (6) can be easily obtained using the calculated $L_{\hat{\boldsymbol{\theta}}}$.

Remark 1 If $X \sim f(\cdot|0, 1)$ then $\sigma X + \mu \sim f(\cdot|\mu, \sigma)$ and we have $h(\sigma X + \mu) = h(X) + \log \sigma = h_o + \log \sigma$ and $VE(\sigma X + \mu) = VE(X) = \nu^2$. Therefore, using (5), we obtain an asymptotic confidence interval for the parameter σ as follows:

$$(\exp\{L_{\hat{\boldsymbol{\theta}}} - h_0 - \frac{Z_{\frac{\alpha}{2}}\nu}{\sqrt{n}}\}, \exp\{L_{\hat{\boldsymbol{\theta}}} - h_0 + \frac{Z_{\frac{\alpha}{2}}\nu}{\sqrt{n}}\}).$$

It is to mention that h_0 and ν^2 are the entropy and varentropy of X, respectively, when $\mu = 0$ and $\sigma = 1$.

One of the most important distributions in the location-scale family is the Laplace distribution. In applied works, the Laplace distribution gives useful representations of many physical situations. Also recently, interest in the Laplace distribution has grown due to its potential transforming application in financial functions. For further study, we can refer to the articles of researchers such as Gel (2010), Baten and Kamil (2009) and so on. The interval estimation of the parameters of this distribution can be seen in articles such as Petropoulos (2011), Alrasheedi (2012), and Jiang and Wong (2012). Indeed if the location parameter is unknown, then finding an optimal confidence interval for the scale parameter is especially important. When μ is known, finding an exact confidence interval for σ is easy, but if both distribution parameters are unknown, then the exact confidence interval for μ and σ can be obtained via the distributions of the pivotal quantities

$$V_n = \frac{\sum_{i=1}^{n} |X_i - \hat{\mu}_n|}{\sigma} \tag{7}$$

and

$$W_n = \frac{\hat{\mu}_n - \mu}{\sum_{i=1}^{n} |X_i - \hat{\mu}_n|}, \tag{8}$$

were $\hat{\mu}_n$ is the maximum likelihood estimation (MLE) of μ. The distribution of V_n and W_n can be derived exactly for small values of n, but calculations become quite tedious as the value of n increases and for $n \geq 3$, that is very difficult. Therefore, for $n \geq 3$, one can use either asymptotic distribution of V_n and W_n (see Bain and Engelhardt (1973)).

Therefore, it seems necessary to find a simple confidence interval for the scale parameter with a suitable coverage probability in the large samples. In the next example, we obtain an asymptotic confidence interval for the Shannon entropy and also calculate two asymptotic confidence intervals for the scale parameter of the Laplace distribution using (4) and (5). In addition, the coverage probability and the average length of the confidence interval of the level $\alpha = 0.05$ have been simulated. In these simulations, 10^6 repetitions were performed, and the average of the obtained results was recorded.

Example 1 Let $X_1, X_2, \ldots, X_n$ be a random sample of the Laplace distribution with the density function $f(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}$. The MLE estimation of σ (when μ is known) and varentropy of X

are $d = \frac{\sum_{i=1}^{n}|X_i - \mu|}{n}$ and $VE(X) = 1$, respectively. Using (6) in this situation, an asymptotic confidence interval for the Shannon entropy is

$$((\log 2d + 1) - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}, (\log 2d + 1) + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}). \tag{9}$$

We know that if μ is unknown, then $\hat{\mu} = X_{[\frac{n+1}{2}]}$ and we need to use d_0 instead of d, where $d_0 = \frac{\sum_{i=1}^{n}|X_i - \hat{\mu}|}{n}$. On the other hand, using (4), an asymptotic confidence interval of the scale parameter σ (when μ is known) is

$$(\frac{d\sqrt{n}}{\sqrt{n} + Z_{\frac{\alpha}{2}}}, \frac{d\sqrt{n}}{\sqrt{n} - Z_{\frac{\alpha}{2}}}). \tag{10}$$

Also, using (5), another asymptotic confidence interval for σ is

$$(de^{\frac{-Z_{\frac{\alpha}{2}}}{\sqrt{n}}}, de^{\frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}}). \tag{11}$$

It is clear that if μ is unknown, then we use d_0 instead d.

In Tables 2 - 7, using several simulation studies, in the level of $\alpha = 0.05$, we examine the coverage probability and length of confidence intervals for Shannon entropy and scale parameter in the Laplace distribution.

Table 2 Coverage probability and length of the confidence interval for the Shannon entropy of the Laplace distribution (μ known and σ unknown)

	$\mu = 5$ and σ unknown				
n	30	50	70	90	150
Coverage probability	0.947	0.948	0.948	0.950	0.950
Length of confidence interval	0.715	0.554	0.468	0.413	0.320

Table 3 The effect of μ on the results for the Shannon entropy in the Laplace distribution

	$n = 50$, μ known and σ unknown				
μ	3	5	7	9	20
Coverage probability	0.947	0.947	0.948	0.947	0.947
Length of confidence interval	0.554	0.554	0.554	0.554	0.554

Table 4 Coverage probability and length of the confidence interval for the Shannon entropy of Laplace distribution (μ unknown and σ unknown)

	μ unknown and σ unknown				
n	30	50	70	90	150
Coverage probability	0.940	0.944	0.946	0.947	0.948
Length of confidence interval	0.715	0.554	0.468	0.413	0.320

Table 5 Coverage probability and length of the confidence interval of σ for the Laplace distribution (μ known)

	$\mu = 5$ (known)				
n	30	50	70	90	150
Coverage probability using (10)	0.952	0.951	0.951	0.951	0.950
Length of confidence interval	1.642	1.201	0.991	0.863	0.657
Coverage probability using (11)	0.947	0.948	0.949	0.949	0.949
Length of confidence interval	1.463	1.123	0.946	0.832	0.643

Table 6 The effect of μ on results of the confidence interval in the Laplace distribution

	$n = 50$ and μ known				
μ	3	5	7	9	20
Coverage probability using (10)	0.951	0.951	0.951	0.951	0.951
Length of confidence interval	1.200	1.201	1.200	1.201	1.201
Coverage probability using (11)	0.948	0.948	0.948	0.948	0.948
Length of CI	1.122	1.123	1.122	1.122	1.123

Table 7 coverage probability and length of the confidence interval of σ for the Laplace distribution (μ unknown)

	μ unknown				
n	30	50	70	90	150
Coverage probability using (10)	0.954	0.952	0.952	0.951	0.950
Length of confidence interval	1.612	1.188	0.984	0.858	0.655
Coverage probability using (11)	0.941	0.945	0.946	0.947	0.948
Length of confidence interval	1.436	1.111	0.939	0.827	0.641

The results obtained in Tables 2 and 4 show that in both cases where the location parameter is known or unknown, the asymptotic confidence interval of entropy has the appropriate coverage probability and the average length of distance. Also, in Tables 5 and 7, we can see that the asymptotic confidence interval of σ in both cases has a good coverage probability and length of the confidence interval. To investigate the effect of μ-changes on the coverage probability and the average length of the confidence interval, Tables 3 and 6 are presented. These tables show that the confidence intervals proposed for the Shannon entropy and scale parameter in the Laplace distribution are not sensitive to μ changes.

2.1. Real data

In this subsection, to illustrate the application of the introduced confidence interval, we will give an example of real data from the Laplace distribution. Bain and Engelhardt (1973) considered 33 years of flood data from two stations on Fox River, Wisconsin (Table 8). They modeled the data using a Laplace distribution and provided 95% approximate confidence intervals for the location and scale parameters based on the pivotal quantities (7) and (8). Kappenman (1975) analyzed further these data for illustrating his conditional approach. Also, using these data, Childs and Balakrishnan (1996) discussed conditional inference of Laplace parameters under Type-II right censoring and Iliopoulos and Balakrishnan (2011) developed exact distributional results for the distributions of the pivotal quantities based on the MLEs of the location and scale parameters of the Laplace distribution based on general Type-II censored samples. These data have been used by Puig and Stephens (2000) and Choi and Kim (2006) for testing goodness of fit for the Laplace distribution.

Table 8 Data on differences in flood stages for two stations on the Fox River, Wisconsin, for 33 different years

1.96	1.96	3.60	3.80	4.79	5.66	5.76	5.78	6.28
6.30	6.76	7.65	7.84	7.99	8.51	9.18	10.13	10.24
10.25	10.43	11.45	11.48	11.75	11.81	12.34	12.78	13.06
13.29	13.98	14.18	14.40	16.22	17.06			

For comparative purposes, we note that both methods of Bain and Engelhardt (1973) and Kappenman (1975) gave the approximate 95% confidence interval for σ to be $[2.49, 4.97]$, while Childs and Balakrishnan (1996) yielded $[2.73, 4.97]$ and Iliopoulos and Balakrishnan (2011) discovered $[2.49, 4.98]$. It should be noted that the maximum likelihood estimate of σ is $\hat{\sigma} = 3.88$.

Using (10) and (11), we conclude two asymptotic confidence intervals $[2.50, 5.10]$ and $[2.38, 4.72]$ for σ respectively. In the Laplace distribution, because the varentropy of the distribution is 1 and does not need to be estimated, our method is simpler than the other proposed methods for calculating the confidence interval for σ (when the location parameter is unknown) and, as can be seen, $[2.38, 4.72]$ has a shorter length than the other methods.

Remark 2 It should be noted that, in addition to location-scale distributions, the above method can be applied to other distributions as well. In some distributions, varentropy depends only on the shape parameter, and if the shape parameter is known, then it means that the varentropy is known. Consequently, with the same method described above, an asymptotic distance estimator can be distributed for another parameter.

Example 2 Let $X_1, X_2, \ldots, X_n$ be a random sample of the gamma distribution with the density function

$$f(x) = \frac{\beta^k}{\Gamma(k)} x^{k-1} e^{-\beta x} \quad k > 0, \beta > 0, x > 0.$$

The family of $gamma(k_0, \beta)$ distributions (where k_0 is any fixed value) is a form of the scale family. If we use (5), then an asymptotic confidence interval can easily be found for β. We know that the MLE estimation of β is $\hat{\beta} = \frac{k_0}{\bar{X}}$. Therefore

$$\boldsymbol{\theta} = (k, \beta), \quad -\frac{1}{n}\sum_{i=1}^{n} \log f(X_i, \hat{\boldsymbol{\theta}}) = \log\Gamma(k_0) - k_0\log(k_0/\bar{X}) - (k_0 - 1)\overline{\log X} + k_0,$$

$$\text{and} \quad h(X) = k_0 - \log\beta + \log\Gamma(k_0) + (1 - k_0)\psi(k_0), \quad VE(X) = (k_0 - 1)^2\dot{\psi}(k_0) - k_0 + 2,$$

where $\psi(\cdot)$ and $\dot{\psi}(\cdot)$ are digamma and trigamma functions, therefore an asymptotic confidence interval of the Shannon entropy is

$$(k_0 \log\frac{\bar{X}}{k_0} + \log\Gamma(k_0) - (k_0 - 1)\overline{\log X} + k_0 - Z_{\frac{\alpha}{2}}\frac{\nu}{\sqrt{n}},$$
$$k_0 \log\frac{\bar{X}}{k_0} + \log\Gamma(k_0) - (k_0 - 1)\overline{\log X} + k_0 + Z_{\frac{\alpha}{2}}\frac{\nu}{\sqrt{n}})$$

and the confidence interval of β is

$$(\exp\{-Z_{\frac{\alpha}{2}}\frac{\nu}{\sqrt{n}} + k_0\log\frac{k_0}{\bar{X}} + (k_0 - 1)\overline{\log X} + (1 - k_0)\psi(k_0)\},$$
$$\exp\{Z_{\frac{\alpha}{2}}\frac{\nu}{\sqrt{n}} + k_0\log\frac{k_0}{\bar{X}} + (k_0 - 1)\overline{\log X} + (1 - k_0)\psi(k_0)\}). \quad (12)$$

When k is known, the exact confidence interval of β is

$$(\frac{\chi^2_{\frac{\alpha}{2}}(2kn)}{2n\overline{X}}, \frac{\chi^2_{1-\frac{\alpha}{2}}(2kn)}{2n\overline{X}}). \quad (13)$$

In Table 9, simulation results for the Shannon entropy in the level of $\alpha = 0.05$ are presented. In Table 11, we compare our asymptotic confidence interval with the exact confidence interval of the scale parameter when the shape parameter is known. Table 11 shows that the confidence interval proposed for the scale parameter in the gamma distribution, relative to the exact confidence interval, has a good coverage probability. Also, the average length of the confidence interval is very slightly different from the exact confidence interval, and as expected, this difference will be very small with increasing sample size. In Tables 10 and 12, we have examined the effect of k on the obtained results. It is observed that increasing the value of k has no effect on the probability of coverage, but by increasing it, the average length of the confidence interval decreases with a gentle slope.

Table 9 Coverage probability and length of the confidence interval for the Shannon entropy of the gamma distribution (k known)

	$k_0 = 2$ and β unknown				
n	30	50	70	90	150
Coverage probability	0.948	0.949	0.949	0.949	0.950
Length of confidence interval	0.575	0.445	0.376	0.332	0.257

Table 10 The effect of k on the results of confidence internal for the Shannon entropy in the gamma distribution (k known)

	$n = 50$ and β unknown				
$k = k_0$	2	4	6	8	12
Coverage probability	0.949	0.950	0.951	0.951	0.953
Length of confidence interval	0.425	0.412	0.404	0.401	0.397

Table 11 Coverage probability and length of the confidence interval for the β in the gamma distribution (k known)

	$k_0 = 2$ and β unknown				
n	30	50	70	90	150
Coverage probability using (12)	0.948	0.950	0.949	0.949	0.951
Length of confidence interval	0.300	0.228	0.191	0.168	0.130
Coverage probability using exact confidence interval (13)	0.950	0.950	0.949	0.949	0.951
Length of confidence interval	0.257	0.198	0.167	0.147	0.113

Table 12 The effect of k on the results of the confidence interval for the β in the gamma distribution (k known)

$k = k_0$	$n = 50$ and β unknown				
	2	4	6	8	12
Coverage probability using (12)	0.946	0.949	0.951	0.951	0.952
Length of confidence interval	0.228	0.211	0.206	0.204	0.203
Coverage probability using exact confidence interval (13)	0.948	0.951	0.948	0.949	0.949
Length of confidence interval	0.197	0.139	0.113	0.098	0.080

3. Interval Estimation for Measure of Uncertainty of Order Statistics

Researchers such as Wong and Chen (1990) and Ebrahimi et al. (2004) have studied the Shannon entropy of order statistics. Abbasnejad and Arghami (2010) and Baratpour and Khammar (2016) also examined the generalized Renyi and Tsallis entropies for order statistics, respectively. Also, recently, the varentropy of order statistics has been introduced by Maadani et al. (2021). In this section, we will introduce an asymptotic confidence interval for the Shannon entropy based on the CLT and varentropy of ith order statistic.

Suppose that $X_1, X_2, ..., X_n$ are independent and identically distributed observations from the cumulative and density function F and f, respectively. If the arrangement of $X_1, X_2, ..., X_n$ from the smallest to the largest denoted as $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ and $f_{i:n}$ is denoted the density function of the ith order statistic, then

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)}[F(x)]^{i-1}[1-F(x)]^{n-i}f(x), \tag{14}$$

where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \;\; a > 0, \; b > 0.$

The entropy of ith order statistics is defined as follows:

$$h(X_{i:n}) = -E(\log f(X_{i:n})) = -\int_{-\infty}^{\infty} f_{i:n}(x) \log f_{i:n}(x)dx$$

Definition 1 The varentropy of ith order statistic is defined as

$$VE(X_{i:n}) = Var(-\log f(X_{i:n})) = \int_{-\infty}^{\infty} f_{i:n}(x)[\log f_{i:n}(x)]^2 dx - [\int_{-\infty}^{\infty} f_{i:n}(x) \log f_{i:n}(x)dx]^2. \tag{15}$$

In general it is difficult to compute $Var(-\log f(X_{i:n}))$. In the next theorem, the method of Maadani et al. (2021), to calculate this variance, using the moment generating function of the random variable $-\log f(X_{i:n})$ is introduced.

Theorem 1 *Let $X_1, X_2, ..., X_n$ be a random sample of size n from the continuous distribution with density function f and cumulative density function F, and let $X_{i:n}$ denote the ith order statistic. Then the varentropy of $X_{i:n}$ can be expressed as*

$$VE(X_{i:n}) = A_{i:n} + C''_{i:n}(1) - [C'_{i:n}(1)]^2, \tag{16}$$

where $$A_{i:n} = [(n-i)^2\dot{\psi}(n-i+1) - (n-1)^2\dot{\psi}(n+1) + (i-1)^2\dot{\psi}(i)],$$
$$C_{i:n}(t) = E(f^{t-1}(F^{-1}(Z_i))),$$

in which $\dot{\psi}$ is trigamma function and $C'_{i:n}(1)$ and $C''_{i:n}(1)$ are first and second derivative of $C_{i:n}(t)$ with respect to t at $t = 1$, respectively and Z_i has Beta distribution with parameters $t(i-1)+1$ and $t(n-i)+1$.

Proof: See Maadani et al. (2021).

For example, if $X \sim U(a,b)$, then

$$E(f^{t-1}(F^{-1}(Z_i)) = \frac{1}{(b-a)^{t-1}}, \quad so\ C'_{i:n}(1) = \log\frac{1}{b-a} \quad \text{and} \quad C''_{i:n}(1) = (\log\frac{1}{b-a})^2.$$

Therefore

$$VE(X_{i:n}) = (n-i)^2\dot{\psi}(n-i+1) - (n-1)^2\dot{\psi}(n+1) + (i-1)^2\dot{\psi}(i), \tag{17}$$

and

$$VE(X_{1:n}) = VE(X_{n:n}) = \frac{(n-1)^2}{n^2}.$$

As can be seen, in the uniform distribution, $VE(X_{i:n})$ does not depend on the distribution parameters. In the following proposition, we show that the varentropy of ith order statistics in the location and scale family is independent of μ and σ.

Proposition 1 *If X has a distribution of location-scale family, then the varentropy of $X_{i:n}$ is independent of location and scale parameters of X.*

Proof: Suppose that X has density and cumulative distribution functions f and F, respectively. If μ and σ are location and scale parameters of X and if $\mu = 0$ and $\sigma = 1$, then

$$f_{i:n}(x) = \frac{1}{B(i,n-i+1)}[F(x)]^{i-1}[1-F(x)]^{n-i}f(x). \tag{18}$$

Also, suppose that $Y_{i:n} = \sigma X_{i:n} + \mu$, the random variable $Y_{i:n}$, has the following density function:

$$g_{i:n}(y) = \frac{1}{\sigma B(i,n-i+1)}[F(\frac{y-\mu}{\sigma})]^{i-1}[1-F(\frac{y-\mu}{\sigma})]^{n-i}f(\frac{y-\mu}{\sigma}). \tag{19}$$

The varentropy is an invariant measure of linear transformation. Therefore, $VE(\sigma X_{i:n} + \mu) = VE(X_{i:n})$, then $VE(X_{i:n}) = VE(Y_{i:n})$ and the proof is completed.

Suppose that $T_1 = (X^{(1)}_{1:n}, \ldots, X^{(1)}_{n:n})$, $T_2 = (X^{(2)}_{1:n}, \ldots, X^{(2)}_{n:n})$, and $T_m = (X^{(m)}_{1:n}, \ldots, X^{(m)}_{n:n})$ are order random vectors of size n from a distribution with density function $f(x, \boldsymbol{\theta})$. It is clear that the sample $X^{(1)}_{i:n}, X^{(2)}_{i:n}, \ldots, X^{(m)}_{i:n}$ is a random sample of size m from the distribution related to ith order statistics from $f(x, \boldsymbol{\theta})$. In order to simplify the symbols, we represent a component of the above random vector, as X_j. Hence we use X_j instead of $X^{(j)}_{i:n}$.

Similar to (5), we can show that the following pivotal variable is asymptotically distributed as standard normal

$$Z_{\hat{\theta}} = \frac{-\frac{1}{m}\sum_{j=1}^{m}\log f_{i;n}(x_j, \hat{\boldsymbol{\theta}}) - h(X_{i:n})}{\sqrt{\frac{1}{m}\sum_{j=1}^{m} Var(\log f_{i:n}(x_j, \hat{\boldsymbol{\theta}}))}}, \tag{20}$$

where $-\frac{1}{m}\sum_{j=1}^{m}\log f_{i;n}(x_j, \hat{\boldsymbol{\theta}}) - h(X_{i:n}))$ and $\frac{1}{m}\sum_{j=1}^{m} Var(\log f_{i:n}(x_j, \hat{\boldsymbol{\theta}}))$ are empirical estimators of $h(X_{i:n})$ and $VE(X_{i:n})$, respectively. Using Proposition 1 in the location-scale family, the varentropy of order statistics is independent of the distribution parameters, and the varentropy no need to estimate in this distributions. Therefore if $VE(X_{i;n}) = \nu^2$, then the pivotal variable (20) is simplified as follows:

$$Z_{\hat{\theta}} = \frac{L - h(X_{i:n})}{\frac{\nu}{\sqrt{m}}}, \tag{21}$$

where $L = -\frac{1}{m}\sum_{j=1}^{m}\log f_{i;n}(x_j, \hat{\boldsymbol{\theta}})$. Hence (21) is suitable for calculating the asymptotic confidence interval for the Shannon entropy of order statistics.

Example 3 Suppose that $X \sim Exp(\theta)$ with the density function $f(x) = \theta e^{-\theta x}, x > 0, \theta > 0$, and that $X_{i:n}$ is ith order statistics of this distribution. An asymptotic confidence interval for the Shannon information related to $X_{i:n}$ is calculated as follows:

$$f_{i:n}(x) = \frac{\theta}{B(i, n-i+1)}(1 - e^{-\theta x})^{i-1}(e^{-\theta x})^{n-i+1}$$

and

$$\begin{aligned} L &= -\frac{1}{m}\sum_{j=1}^{m} \log f_{i;n}(x_j, \hat{\boldsymbol{\theta}}) \\ &= \log B(i, n-i+1) - \frac{i-1}{m}\sum_{j=1}^{m} \log(1 - e^{-\hat{\theta} x_j}) - \frac{(n-i+1)\hat{\theta}}{m}\sum_{j=1}^{m} \log x_j - \log \hat{\theta}, \end{aligned} \tag{22}$$

where $\hat{\theta}$ is an MLE estimator of θ in the distribution of $X_{i:n}$ that can be calculated from the following equation:

$$(i-1)\sum_{j=1}^{m} \frac{x_j e^{-\theta x_j}}{1 - e^{-\theta x_j}} - (n-i+1)\sum_{j=1}^{m} \log x_j + \frac{m}{\theta} = 0.$$

On the other hand using Theorem 1, we can see that if $X \sim Exp(\theta)$, then the varentropy of ith order statistics is

$$VE(X_{i:n}) = (n-i)^2\dot{\psi}(n-i+1) - (n-1)^2\dot{\psi}(n+1) + (i-1)^2\dot{\psi}(i) = \upsilon^2$$

$$VE(X_{1:n}) = 1 \quad \text{and} \quad VE(X_{n:n}) = 1 + \frac{\pi^2}{6} + (1-2n)\dot{\psi}(i).$$

We see that the $VE(X_{i:n})$ is independent of θ and (21) implies

$$h(X_{i:n}) \in (\log B(i, n-i+1) - \frac{i-1}{m}\sum_{j=1}^{m} \log(1 - e^{-\hat{\theta} x_j}) - \frac{(n-i+1)\hat{\theta}}{m}\sum_{j=1}^{m} \log x_j - \log \hat{\theta} \mp \frac{Z_{\frac{\alpha}{2}}\upsilon}{\sqrt{m}}). \tag{23}$$

Also,

$$h(X_{1:n}) \in (-\log n - \frac{n\hat{\theta}}{m}\sum_{j=1}^{m} \log x_j - \log \hat{\theta} \mp \frac{Z_{\frac{\alpha}{2}}}{\sqrt{m}})$$

and

$$h(X_{n:n}) \in (-\log n - \frac{n-1}{m}\sum_{j=1}^{m} \log(1 - e^{-\hat{\theta} x_j}) - \frac{\hat{\theta}}{m}\sum_{j=1}^{m} \log x_j - \log \hat{\theta} \mp \frac{Z_{\frac{\alpha}{2}}[1 + \frac{\pi^2}{6} + (1-2n)\dot{\psi}(i)]}{\sqrt{m}}).$$

Table 13 Coverage probability and length of the confidence interval of the $h(X_{i:n})$ from the exponential distribution

	$\theta = 1$	$n = 20$	$m = 30$		
i	1	5	10	15	20
Coverage probability	0.9464	0.9523	0.9542	0.9556	0.9535
Length of confidence interval	0.7156	0.5274	0.5186	0.5216	0.5749

Table 14 Coverage probability and length of the confidence interval of the $h(X_{i:n})$ from the exponential distribution

	$\theta=1$	$n=20$	$m=40$		
i	1	5	10	15	20
Coverage probability	0.9479	0.9502	0.9534	0.9541	0.9533
Length of confidence interval	0.6197	0.4568	0.4491	0.4517	0.4979

Table 15 Coverage probability and length of the confidence interval of the $h(X_{i:n})$ from the exponential distribution

	$\theta=1$	$n=20$	$m=100$		
i	1	5	10	15	20
Coverage probability	0.9501	0.9498	0.9510	0.9522	0.9512
Length of CI	0.3919	0.2889	0.2840	0.2865	0.3149

Using the R software and simulation method, we generated data with volumes of 30, 40, and 100 from the distribution of ith order statistic when i is 1,5,10,15,20 from the order sample of size $n = 20$ of the standard exponential distribution. Then we used (23) to find an asymptotic confidence interval of Shannon information. The coverage probability and the average length of confidence interval were recorded in Tables 13-15 for 10^6 repetitions. Using (21), the length of the confidence interval for the Shannon entropy of $X_{i:n}$ is $\frac{2\nu Z_{\frac{\alpha}{2}}}{\sqrt{m}}$. Since $Z_{\frac{\alpha}{2}}$ and m are constant in the certain sample and known level of α, the length of confidence interval depends only on the varentropy of $X_{i:n}$. Maadani et al. (2021) showed that in the exponential distribution, the value of the $VE(X_{i:n})$ is symmetric with respect to i and around the median of the order statistics, we have a minimum varentropy. Therefore, the interval estimate of $h(X_{i:n})$ has the minimum length around the median of order statistics. The tables 13-15 confirm this result.

4. Hypothesis Testing

In the family of location-scale distributions, the Shannon entropy depends only on the scale parameter. If we denote the scale parameter of the distribution by θ, then the Shannon entropy is a function of θ, and $h(X) = g(\theta)$. Now if we have the hypothesis testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, we can replace the $H_0 : g(\theta) = g(\theta_0)$ versus $H_1 : g(\theta) \neq g(\theta_0)$ instead. So using (5), we get the following test statistic:

$$Z_{\hat{\theta}} = \frac{-\frac{1}{n}\sum_{i=1}^{n} \log f(X_i, \hat{\theta}) - g(\theta_0)}{\sqrt{\frac{VE(X)}{n}}}. \tag{24}$$

Therefore using (24), if $|Z_{\hat{\theta}}| > Z_{\frac{\alpha}{2}}$, then we can reject H_0 in the level α. It is easy to see that if H_0 is true, then $g(\theta) = g(\theta_0)$, $Z_{\hat{\theta}}$ goes to the standard normal and type (I) error probability is

$$P(|Z_{\hat{\theta}}| > Z_{\frac{\alpha}{2}}) \to P(|Z| > Z_{\frac{\alpha}{2}}) = \alpha.$$

Moreover if $g(\theta_0) \neq g(\theta)$, then $Z_{\hat{\theta}}$ goes to $-\infty$ or $+\infty$ as $n \to \infty$. Therefore, the power of testing goes to 1.

Example 4 The below data shows 25 observations of gamma distribution with density function $f(x) = \frac{\beta^k}{\Gamma(k)} x^{k-1} e^{-\beta x} \quad k > 0, \beta > 0, x > 0$, with parameters $(2, 3)$.

Table 16 Data to test $H_0 : \beta = 3$ versus $H_1 : \beta \neq 3$

0.6099	0.8280	0.3148	0.4960	1.0615	0.5106	0.4145	0.3023	1.6312
0.1825	0.5679	0.6105	2.2167	0.5312	0.6805	0.9944	0.4933	0.2759
0.5485	0.1700	0.3192	0.9597	0.2948	0.9766	0.3782		

Using (24) the test statistic for this example is

$$Z_{\hat{\theta}} = \frac{2log\frac{\bar{X}}{2} - \overline{\log X} + \log\beta + \psi(2)}{\sqrt{\frac{\dot{\psi}(2)}{n}}} \quad \bar{X} = 0.6947 \quad \overline{\log X} = -0.5791 \quad then \quad Z_{\hat{\theta}} = -0.089$$

We can see that $|Z_{\hat{\theta}}| < 1.96$. Therefore, in the level of $\alpha = 0.05$, there is no reason to reject the null hypothesis.

Similar to what was said, if we want to test the hypothesis $H_0 : h(X_{i:n}) = h_0$ versus $H_1 : h(X_{i:n}) \neq h_0$, we can use of test statistics (20).

Example 5 Suppose that $X \sim Exp(1)$ and that $X_{i:n}$ is the ith order statistics in this distribution. Using R software, we generate the 30 random samples from the distribution of $X_{3:20}$. We are interested in testing hypothesis $H_0 : h(X_{3:20}) \leq -2$ versus $H_1 : h(X_{3:20}) > -2$. It is to mention that the Shannon information of $X_{3:20}$ in this assumption is -1.09.

Data:

0.2660	0.2086	0.2459	0.1035	0.1230	0.1441	0.1632	0.1671	0.2961
0.1366	0.1116	0.0240	0.1611	0.0680	0.2739	0.1347	0.1931	0.2140
0.0693	0.4196	0.2748	0.0533	0.0646	0.1247	0.0900	0.0800	0.2882
0.2358	0.0870	0.2188						

We have $m = 30$. Using Theorem 1 we obtain $\nu = 0.762$ and $L = -\frac{1}{m}\sum_{j=1}^{m} \log f_{i;n}(x_j, \hat{\boldsymbol{\theta}}) = -1.055$. Therefore, the test statistics (21) implies $Z_{\hat{\theta}} = 6.7$ and $Z_{\hat{\theta}} > 1.645$. Then the null hypothesis is rejected.

5. Conclusions

In the location-scale family, estimating the scale parameter in some distributions such as the Laplace distribution, is of particular importance when the location parameter is unknown. Also estimating the Shannon entropy for the ith order statistics, when the distribution parameters are unknown, is very important in information theory and engineering sciences. Similar to what is done in the interval estimating to the mean in large samples, we introduced an asymptotic confidence interval for the scale parameter of distributions and the Shannon information of the ith order statistics in this family. We showed that similar to $Var(-\log f(X))$, $Var(-\log f(X_{i:n}))$ is independent of location and scale parameters, so we have a known variance and there will be no variance estimation error. Therefore, this will help to get the optimal estimate. Using several simulation studies, we showed that those confidence intervals have a good coverage probability and length. Also, for the the scale parameter of Laplace distribution, using real data, we compared our proposed confidence interval with the confidence intervals of other researchers. In this comparison, it was found that the obtained confidence interval, seems appropriate due to the simplicity of calculation and shorter length. Finally, using the presented pivotal quantity, we performed two tests of the statistical hypothesis and examined its results.

Acknowledgments

The authors would like to thank the editor and anonymous referees for their valuable comments and suggestions that improved the quality of the paper.

References

Abbasnejad M, Arghami NR. Renyi entropy properties of order statistics. Comm Stat Theo Meth. 2010; 40(1): 40-52.

Afhami B, Madadi M, Rezapour M. Goodness-of-fit test based on Shannon entropy of k-record value from the generalized Pareto distribution. Stat Sci. 2015; 9(1): 43-60

Alizadeh Noughabi H, Arghami NR. Goodness-of-fit tests based on correcting moments of entropy estimators. Comm Stat Simul Comput. 2013; 42(3): 499-513.

Alrasheedi M. Confidence intervals for double exponential distribution: A Simulation Approach. Int J Phys Math Sci. 2012; 6(1): 84-88

Arikan E. Varentropy decreases under the polar transform. IEEE Trans Inf Theo. 2016; 62(6): 3390-3400.

Bain LJ, Engelhardt M. Interval estimation for the two parameter double exponential distribution. Technometrics. 1973; 15(4): 875- 887.

Baratpour S, Khammar A. Tsallis entropy properties of order statistics and some stochastic comparisons. JSRI. 2016; 13(1): 25-41.

Baten A, Kamil A. Inventory management systems with hazardous items of two-parameter exponential distribution. J Sos Sci. 2009; 5(3): 183-187.

Bobkov S, Madiman M. Concentration of the information in data with log-concave distributions. Ann Probab. 2011; 39(4): 1528-1543.

Casella G, Berger RL. Statistical inference. Belmont California: Wadsoworth Publishing; 1990.

Childs A, Balakrishnan N. Conditional inference procedures for the Laplace distribution based on Type-II right censored samples. Stat Probab Lett. 1996; 31(1): 31-39.

Choi B, Kim K. Testing goodness-of-fit for Laplace distribution based on maximum entropy. Statistics. 2006; 40(6): 517-531.

Di Crescenzo A, Paolillo L. Analysis and applications of the residual varentropy of random lifetimes. Probab Eng Inf Sci. 2020; 35(3): 680-698.

Ebrahimi N, Soofi ES, Zahedi H. Information properties of order statistics and spacings. IEEE Trans Inf Theo. 2004; 50(1): 177-183.

Enomoto R, Okamoto N, Seo T. On the asymptotic normality of test statistics using Songs kurtosis. J Stat Theo Pract. 2013; 7(1): 102-119.

Gel YR. Test of fit for a Laplace distribution against heavier tailed alternatives. Comput Stat Data Anal. 2010; 54(4): 958-965.

Iliopoulos G, Balakrishnan N. Exact likelihood inference for Laplace distribution based on Type-II censored samples. J Stat Plan Inf. 2011; 141(3): 1224-1239.

Jager L, Wellner JA. Goodness-of-fit Tests via Phi-divergences. Ann Stat. 2007; 35(5): 20182053.

Jiang L, Wong A. Interval estimations of the two-parameter exponential distribution. J Probab Stat. 2012; doi:10.1155/2012/734575.

Kappenman RF. Conditional confidence intervals for double exponential distribution parameters. Technometrics.1975; 17(2): 233-235.

Kontoyannis I, Verdu S. Optimal lossless compression: Source varentropy and dispersion. IEEE International Symposium Information Theory. 2013 Jul 7-12; Turkey. Istanbul: IEEE Xplore; 2013. p. 1739-1743.

Maadani S, Mohtashami Borzadaran GR, Rezaei Roknabadi AH. A new generalized varentropy and its properties. Ural math J. 2020; 6(1): 114-129.

Maadani S, Mohtashami Borzadaran GR, Rezaei Roknabadi AH. Varentropy of order statistics and some stochastic comparisons. Comm Stat Theo Meth. 2021; doi:10.1080/03610926.2020.1861299.

Menendez ML. Shannons entropy in exponential families, statistical applications. Appl Math Lett. 2000; 13(1): 37-42.

Pasha E, Khodabin M, Mohtashami Borzadaran GR . Testing statistical hypothesis via Shannon's entropy in exponential families. Iranian Int J Sci. 2004; 5(2): 267-279

Petropoulos C. New classes of improved confidence intervals for the scale parameter of a two parameter exponential distribution. Stat Methodol. 2011; 8(4): 401-410.

Puig P, Stephens MA. Tests of fit for the Laplace distribution, with applications. Technometrics. 2000; 42(4): 417-424.

Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948; 27(3): 379-423.

Song KS. Renyi information, log likelihood and an intrinsic distribution measure. J Stat Plan Infer. 2001; 93(1-2): 51-69.

Taufer E. On entropy based tests for exponentiality. Commu Stat Simul Comput. 2002; 31(2): 189-200.

Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. J Stat Phys. 1988; 52(1): 479-487.

Wong KM, Chen S. The entropy of ordered sequences and order statistics. IEEE Trans Inf Theo. 1990; 36(2): 276-284.

Zamanzade E, Arghami NR. Testing normality based on new entropy estimators. J Stat Comput Simul. 2012; 82(11): 1701-1713.

Zografos K. On Mardias and Songs measures of kurtosis in elliptical distributions. J Multivar Anal. 2008; 99(5): 858-879.